SEQUENCE LISTING

<110> McIninch, James

<120> COMPUTATIONAL NUCLEIC ACID CODING AND FEATURE ANALYSIS

<130> 04983.0220.00US00

<160> 4

<170> PatentIn version 3.0

<210> 1
<211> 2165
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> unsure
<222> (1)...(2165)
<223> Unsure at all n locations

<220>
<223> Ecotype Landsberg, genomic DNA

<400> 1
tactcaaaaa tatattccat gcttaattag gccggattcg cggtgacgat gcaccaagag 60

cggtttttcc gagcattgta ggccgtcctc gccacaccgg tgtgatggtt gggatgggac 120

aaaaggatgc ttatgttgga gacgaggctc aatcaaaacg tggtatcttg actctgaagt 180

acccaattga gcatggaatt gttaataatt gggatgacat ggagaagatt tggcatcaca 240

ctttctacaa tgagcttcgt gttgcccctg aagaacatcc ggttctcttg accgaagctc 300

ctctcaatcc gaaagctaac cgtgagaaga tgactcagat catgtttgag acattcaata 360

ctcctgctat gtatgttgcc attcaagctg ttctctcact ctatgccagt ggccgtacta 420

ctggtcagta cattactaca ttctttttat accgtttggt tgaaataaaa ttcggtttgg 480

ttcgattcga gtttgctctc attatttta ttttgttggt taggtattgt tttggactcc 540

ggagatggtg tgagccacac ggtaccaatc tacgagggtt atgcacttcc acacgcaatc 600

ctgcgtcttg atcttgcagg tcgtgaccta accgaccacc ttatgaaaat cctgacagag 660

cgtggttact ctttcaccac aactgctgag cgtgagattg ttagagacat gaaggagaag 720

ctctcttaca ttgccttgga ctttgaacaa gagctcgaga cttccaaaac aagctcatcc 780

gttgagaaga gcttcgagct gccagacggt caagtgatca ccatcggggc agagcgtttc 840

cgatgccctg aagttctgtt tcagccatcg atgatcggaa tggaaaatcc gggaattcat 900

gaaactactt acaactcaat catgaaatgt gatgtggata tcaggaagga tctttatgga 960

aacattgtgc ttagtggtgg caccacaatg ttcgatggga ttggtgatag gatgagtaaa 1020

gagatcacag cgttggctcc aagcagtatg aacatcaaag tggtggctcc accggaaagg 1080

aagtacagtg tctggatcgg tggctctatc ttggcttccc tcagtacttt ccagcaggta 1140

aattacttac tatacttaat acataaagtc tattagtgat ttgatgtata aagtgttaca 1200

aaaatgtgtt ccaaatttgc agatgtggat tgcgaaagcg gagtatgatg aatctggacc 1260

gtcaatcgtc cacaggaagt gcttctgatc aaaagtcacc aagtaaaaca agagcggtaa 1320

aaattttgat atcagttttt caccctgaag ccagttgcta taattactca caacttctct 1380

atttgtgttc ttttattctt gtccctcgtt gttcatttta atctcttttt tgcaacaaag 1440

caacttaaaa aaacagagca gtcattaaca gaatgttatt attatatata tgtatacata 1500

ttagtataca cccattattt cattaaaaca tttatcatat aaggatagga ttctatacat 1560

cgatatattt attttgttga cactattcag cacatgctta tgtcttatct tgttagtata 1620

tgtaaccaaa gacaaataat agatgctaca aattgttttc tttgaagcaa aaatttcaat 1680

cttaaaattg ttttttttcca ggttacacaa aaaaaacttg tagtttgtaa attttctata 1740

caattttggg gatctcaaca agaacatgaa cttcaacttc tagtcatatg acgacctgag 1800

tctgcgcggc tgtgaatctc tttgctgcag taaatgttta caagtggtgt gtaaattggt 1860

actgattcaa aagctttaag aaatctacac atttcgtgaa attatttagc agacttgata 1920

ttaaaaatct aggataaaat gactatccaa agacaaatag gactgtttca catgttcccc 1980

tgattcttgt agctcataac tcatcagcag ttaacttttc tacctcatac acgctcgcaa 2040

tncgtttgga attatcagct ntaatttttc taattctttg gaaattatta gcagctcgat 2100

caaatggggc atggcttctt cttctatctg caactcatct aaactttcca tgaagaaaca 2160

aagct 2165

<210> 2
<211> 423
<212> PRT
<213> Unknown

<220>
<223> Describes a predicted protein sequence

<220>
<221> site
<222> (1)...(423)
<223> A stop codon is predicted at all XAA locations

<400> 2

Xaa Arg Phe Phe Arg Ala Leu Xaa Ala Val Leu Ala Thr Pro Val Xaa
1 5 10 15

Trp Leu Gly Trp Asp Lys Arg Met Leu Met Leu Glu Thr Arg Leu Asn
20 25 30

Gln Asn Val Val Ser Xaa Leu Xaa Ser Thr Gln Leu Ser Met Glu Leu
35 40 45

Leu Ile Ile Gly Met Thr Trp Arg Arg Phe Gly Ile Thr Leu Ser Thr
50 55 60

Met Ser Phe Val Leu Pro Leu Lys Asn Ile Arg Xaa Leu Thr Glu Ala
65 70 75 80

Pro Leu Asn Pro Lys Ala Asn Arg Glu Lys Met Thr Gln Ile Met Phe
85 90 95

Glu Thr Phe Asn Thr Pro Ala Met Tyr Val Ala Ile Gln Ala Val Leu
100 105 110

Ser Leu Tyr Ala Ser Gly Arg Thr Thr Gly Gln Tyr Ile Thr Thr Phe
115 120 125

Phe Leu Tyr Arg Xaa Ser Gly Asp Gly Val Ser His Thr Val Pro Ile
130 135 140

Tyr Glu Gly Tyr Ala Leu Pro His Ala Ile Leu Arg Leu Asp Leu Ala
145 150 155 160

Gly Arg Asp Leu Thr Asp His Leu Met Lys Ile Leu Thr Glu Arg Gly
165 170 175

Tyr Ser Phe Thr Thr Thr Ala Glu Arg Glu Ile Val Arg Asp Met Lys
180 185 190

Glu Lys Leu Ser Tyr Ile Ala Leu Asp Phe Glu Gln Glu Leu Glu Thr
195 200 205

Ser Lys Thr Ser Ser Ser Val Glu Lys Ser Phe Glu Leu Pro Asp Gly
210 215 220

Gln Val Ile Thr Ile Gly Ala Glu Arg Phe Arg Cys Pro Glu Val Leu
225 230 235 240

Phe Gln Pro Ser Met Ile Gly Met Glu Asn Pro Gly Ile His Glu Thr
245 250 255

Thr Tyr Asn Ser Ile Met Lys Cys Asp Val Asp Ile Arg Lys Asp Leu
260 265 270

Tyr Gly Asn Ile Val Leu Ser Gly Gly Thr Thr Met Phe Asp Gly Ile
275 280 285

Gly Asp Arg Met Ser Lys Glu Ile Thr Ala Leu Ala Pro Ser Ser Met
290 295 300

Lys Ile Lys Val Val Ala Pro Pro Glu Arg Lys Tyr Ser Val Trp Ile
305 310 315 320

Gly Gly Ser Ile Xaa Val Pro Asn Leu Gln Met Trp Ile Ala Lys Ala
325 330 335

Glu Tyr Xaa Asn Leu Asp Arg Gln Ser Ser Thr Gly Ser Ala Ser Asp
340 345 350

Gln Lys Ser Pro Ser Lys Thr Arg Ala Val Lys Ile Leu Xaa Asn Ser
355 360 365

Ser Ala Val Asn Phe Ser Thr Ser Tyr Thr Leu Ala Ile Arg Leu Glu
370 375 380

Leu Ser Ala Leu Ile Phe Leu Ile Ser Leu Glu Ile Ile Ser Ser Ser
385 390 395 400

Ile Lys Trp Gly Met Ala Ser Ser Ser Ile Cys Asn Ser Ser Lys Leu
405 410 415

Ser Met Lys Lys Gln Ser Xaa
420

<210> 3
<211> 422
<212> PRT
<213> Unknown

<220>
<223> Describes a predicted protein sequence

<220>
<221> site
<222> (1)...(422)
<223> A stop codon is predicted at all XAA locations

<400> 3

Xaa Arg Phe Phe Arg Ala Leu Xaa Ala Val Leu Ala Thr Pro Val Xaa
1 5 10 15

Trp Leu Gly Trp Asp Lys Arg Met Leu Met Leu Glu Thr Arg Leu Asn

20 25 30

Gln Asn Val Val Ser Xaa Leu Xaa Ser Thr Gln Leu Ser Met Glu Leu
35 40 45

Leu Ile Ile Gly Met Thr Trp Arg Arg Phe Gly Ile Thr Leu Ser Thr
50 55 60

Met Ser Phe Val Leu Pro Leu Lys Asn Ile Arg Xaa Leu Thr Glu Ala
65 70 75 80

Pro Leu Asn Pro Lys Ala Asn Arg Glu Lys Met Thr Gln Ile Met Phe
85 90 95

Glu Thr Phe Asn Thr Pro Ala Met Tyr Val Ala Ile Gln Ala Val Leu
100 105 110

Ser Leu Tyr Ala Ser Gly Arg Thr Thr Gly Gln Tyr Ile Thr Thr Phe
115 120 125

Phe Leu Tyr Arg Xaa Ser Gly Asp Gly Val Ser His Thr Val Pro Ile
130 135 140

Tyr Glu Gly Tyr Ala Leu Pro His Ala Ile Leu Arg Leu Asp Leu Ala
145 150 155 160

Gly Arg Asp Leu Thr Asp His Leu Met Lys Ile Leu Thr Glu Arg Gly
165 170 175

Tyr Ser Phe Thr Thr Thr Ala Glu Arg Glu Ile Val Arg Asp Met Lys
180 185 190

Glu Lys Leu Ser Tyr Ile Ala Leu Asp Phe Glu Gln Glu Leu Glu Thr
195 200 205

Ser Lys Thr Ser Ser Ser Val Glu Lys Ser Phe Glu Leu Pro Asp Gly
210 215 220

Gln Val Ile Thr Ile Gly Ala Glu Arg Phe Arg Cys Pro Glu Val Leu
225 230 235 240

Phe Gln Pro Ser Met Ile Gly Met Glu Asn Pro Gly Ile His Glu Thr
245 250 255

Thr Tyr Asn Ser Ile Met Lys Cys Asp Val Asp Ile Arg Lys Asp Leu
260 265 270

Tyr Gly Asn Ile Val Leu Ser Gly Gly Thr Thr Met Phe Asp Gly Ile
275 280 285

Gly Asp Arg Met Ser Lys Glu Ile Thr Ala Leu Ala Pro Ser Ser Met
290 295 300

Lys Ile Lys Val Val Ala Pro Pro Glu Arg Lys Tyr Ser Val Trp Ile

305 310 315 320

Gly Gly Ser Ile Leu Ala Ser Xaa Gln Met Trp Ile Ala Lys Ala Glu
325 330 335

Tyr Xaa Asn Leu Asp Arg Gln Ser Ser Thr Gly Ser Ala Ser Asp Gln
340 345 350

Lys Ser Pro Ser Lys Thr Arg Ala Val Lys Ile Leu Xaa Asn Ser Ser
355 360 365

Ala Val Asn Phe Ser Thr Ser Tyr Thr Leu Ala Ile Arg Leu Glu Leu
370 375 380

Ser Ala Leu Ile Phe Leu Ile Ser Leu Glu Ile Ile Ser Ser Ser Ile
385 390 395 400

Lys Trp Gly Met Ala Ser Ser Ser Ile Cys Asn Ser Ser Lys Leu Ser
405 410 415

Met Lys Lys Gln Ser Xaa
420

<210> 4
<211> 296
<212> PRT
<213> Arabidopsis thaliana

<220>
<223> Ecotype columbia, describes actin

<400> 4

Met Glu Lys Ile Trp His His Thr Phe Tyr Asn Glu Leu Arg Val Ala
1 5 10 15

Pro Glu Glu His Pro Val Leu Leu Thr Glu Ala Pro Leu Asn Pro Lys
20 25 30

Ala Asn Arg Glu Lys Met Thr Gln Ile Met Phe Glu Thr Phe Asn Thr
35 40 45

Pro Ala Met Tyr Val Ala Ile Gln Ala Val Leu Ser Leu Ala Ser Gly
50 55 60

Arg Thr Thr Gly Gly Ile Val Leu Asp Ser Gly Asp Gly Val Ser His
65 70 75 80

Thr Val Pro Ile Tyr Glu Gly Tyr Ala Leu Pro His Ala Ile Leu Arg
85 90 95

Leu Asp Leu Ala Gly Arg Asp Leu Thr Asp His Leu Met Lys Ile Leu
100 105 110

```
Thr Glu Arg Gly Tyr Ser Phe Thr Thr Thr Ala Glu Arg Glu Ile Val
        115                 120                 125

Arg Asp Met Lys Glu Lys Leu Ser Tyr Ile Ala Leu Asp Phe Glu Gln
    130                 135                 140

Glu Leu Glu Thr Ser Lys Thr Ser Ser Ser Val Glu Lys Ser Phe Glu
145                 150                 155                 160

Leu Pro Asp Gly Gln Val Ile Thr Ile Gly Ala Glu Arg Phe Arg Cys
                165                 170                 175

Pro Glu Val Leu Phe Gln Pro Ser Met Ile Gly Met Glu Asn Pro Gly
            180                 185                 190

Ile His Glu Thr Thr Tyr Asn Ser Ile Met Lys Cys Asp Val Asp Ile
        195                 200                 205

Arg Lys Asp Leu Tyr Gly Asn Ile Val Leu Ser Gly Gly Thr Thr Met
    210                 215                 220

Phe Gly Gly Ile Gly Asp Arg Met Ser Lys Glu Ile Thr Ala Leu Ala
225                 230                 235                 240

Pro Ser Ser Met Lys Ile Lys Val Val Ala Pro Pro Glu Arg Lys Tyr
                245                 250                 255

Ser Val Trp Ile Gly Gly Ser Ile Leu Ala Ser Leu Ser Thr Phe Gln
            260                 265                 270

Gln Met Gln Met Trp Ile Ala Lys Ala Glu Tyr Asp Glu Ser Gly Pro
        275                 280                 285

Ser Ile Val His Arg Lys Cys Phe
    290                 295
```